# 最小二乗法と単回帰分析

本稿は、不動産鑑定士の有志で行なっていた不動産金融工学勉強会において筆者が提出したレジュメに、今回若干加筆したものである。

ヘドニックアプローチによる不動産価格や賃料の推定、取引事例比較法に活用する比準表を作成するための要因格差率の推定、収益還元法に用いる割引率や還元利回りを事例から推定する場合など、あらゆる場面で回帰分析が有用である。そのベースとなる最小二乗法のカラクリを、数式を丁寧に解くことにより説明する。

なお、回帰分析等の多変量解析は、テクニックとして金融工学にも有用であるのであって、従来から様々な分野で広く用いられているものである。

---

【最小二乗法による回帰式の推定】

$$\hat{y} = \hat{\alpha} + \hat{\beta}\, x \qquad \cdot\cdot\cdot [1]$$

但し、^のつくものは予測値であることを表す。

上記[1]式は、y と x の間に何らかの関係があることを前提に、y を x で説明しようとするものである。y を目的変数（従属変数）といい、x を説明変数（独立変数）という。

実測値 $y_i$ と予測値 $\hat{y}_i$ との差 $e_i$ を残差という。

$$y_i = \hat{\alpha} + \hat{\beta}\, x_i + e_i \qquad \cdot\cdot\cdot [2]$$

したがって、

$$e_i = y_i - \hat{y}_i$$

$$= y_i - (\hat{\alpha} + \hat{\beta}\, x_i) \qquad \cdot\cdot\cdot [3]$$

残差 $e_i$ の平方和（二乗和）を $S_e$ とし、下式の $S_e$ が最小となるような $\hat{\alpha}$ と $\hat{\beta}$ を求めるのが、最小二乗法である。

$$S_e = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \{y_i - (\hat{\alpha} + \hat{\beta}\, x_i)\}^2 \qquad \cdot\cdot\cdot [4]$$

そのために$S_e$を$\hat{\alpha}$と$\hat{\beta}$でそれぞれ偏微分してゼロとおき連立方程式を解く。

$$\begin{cases} \dfrac{\partial S_e}{\partial \hat{\alpha}} = -2\sum_{i=1}^{n}\{y_i - (\hat{\alpha} + \hat{\beta}\, x_i)\} = 0 & \cdots [5] \\ \\ \dfrac{\partial S_e}{\partial \hat{\beta}} = -2\sum_{i=1}^{n} x_i\{y_i - (\hat{\alpha} + \hat{\beta}\, x_i)\} = 0 & \cdots [6] \end{cases}$$

[5]式、[6]式を整理すると、

正規方程式

$$\boxed{\begin{array}{ll} \sum y_i = \hat{\alpha}\, n + \hat{\beta}\sum x_i & \cdots [7] \\ \\ \sum x_i y_i = \hat{\alpha}\sum x_i + \hat{\beta}\sum {x_i}^2 & \cdots [8] \end{array}}$$

が得られる。

[7]式より、

$$\hat{\alpha} = \frac{\sum y_i}{n} - \hat{\beta}\frac{\sum x_i}{n} = \overline{y} - \hat{\beta}\,\overline{x} \qquad \cdots [9]$$

但し、$\overline{\phantom{x}}$のつくものは、算術平均を表す。

[9]式を[8]式に代入すると、

$$\sum x_i y_i = \left(\frac{\sum y_i}{n} - \hat{\beta}\frac{\sum x_i}{n}\right)\sum x_i + \hat{\beta}\sum {x_i}^2$$

$$\hat{\beta}\left\{\sum {x_i}^2 - \frac{\left(\sum x_i\right)^2}{n}\right\} = \sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{n}$$

$$\therefore \hat{\beta} = \frac{\sum x_i y_i - \dfrac{\sum x_i \cdot \sum y_i}{n}}{\sum {x_i}^2 - \dfrac{\left(\sum x_i\right)^2}{n}} \qquad \cdots [10]$$

ここで、[10]式右辺分子を、

$$\sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{n} = \sum\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) = S_{xy}$$ とし、

[10]式右辺分母を、

$$\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n} = \sum\left(x_i - \bar{x}\right)^2 = S_{xx}$$ とすれば、$\hat{\beta}$は、

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \qquad ・・・[11]$$

と表現できる。$S_{xy}$は x と y の偏差積和（各々の平均からの隔たりを掛けたものの和）、$S_{xx}$は x の偏差平方和（各 x の平均からの隔たりを２乗したものの和）である。

以上より、次の単回帰式の推定式を得る。

$$\hat{y} = \hat{\alpha} + \hat{\beta}\,x = \bar{y} + \frac{S_{xy}}{S_{xx}}\left(x - \bar{x}\right) \qquad ・・・[12]$$

$\because$ [9]式 $\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x}$ を、[1]式 $\hat{y} = \hat{\alpha} + \hat{\beta}\,x$ に代入すれば、

$\hat{y} = \bar{y} - \hat{\beta}\,\bar{x} + \hat{\beta}\,x = \bar{y} + \hat{\beta}\left(x - \bar{x}\right)$ となるからである。

---

次に、x と y の関係を示す実際のデータ２組$\left(x_1, y_1\right)\left(x_2, y_2\right)$があるものとして、これらから具体的に式を展開する。無論、たった２つの標本から母集団を推定することはできないが、上記一般式が正しいことを直感的に理解するためである。

$$\begin{cases} y_1 = \alpha + \beta\,x_1 + e_1 \\ y_2 = \alpha + \beta\,x_2 + e_2 \end{cases}$$

残差平方和$S_e$は、

$$S_e = e_1^{\,2} + e_2^{\,2}$$

$$= \left\{y_1 - \left(\alpha + \beta\,x_1\right)\right\}^2 + \left\{y_2 - \left(\alpha + \beta\,x_2\right)\right\}^2$$

$$= y_1^{\,2} - 2y_1\left(\alpha + \beta\,x_1\right) + \left(\alpha + \beta\,x_1\right)^2 + y_2^{\,2} - 2y_2\left(\alpha + \beta\,x_2\right) + \left(\alpha + \beta\,x_2\right)^2$$

$$= y_1^{\,2} - 2y_1\left(\alpha + \beta\,x_1\right) + \alpha^2 + 2\alpha\,\beta\,x_1 + \beta^2 x_1^{\,2}$$
$$+ y_2^{\,2} - 2y_2\left(\alpha + \beta\,x_2\right) + \alpha^2 + 2\alpha\,\beta\,x_2 + \beta^2 x_2^{\,2}$$

$S_e$ を α と β でそれぞれ偏微分してゼロとおく。

$$\frac{\partial S_e}{\partial \alpha} = -2y_1 + 2\alpha + 2\beta x_1 - 2y_2 + 2\alpha + 2\beta x_2$$
$$= -2(y_1 + y_2 - \beta x_1 - \beta x_2 - 2\alpha) = 0$$

$$\therefore y_1 + y_2 = 2\alpha + \beta(x_1 + x_2) \quad \cdot\cdot\cdot [13]$$

→この[13]式は、前記正規方程式の[7]式 $\sum y = \alpha n + \beta \sum x$ に等しい。

$$\frac{\partial S_e}{\partial \beta} = -2x_1y_1 + 2\alpha x_1 + 2\beta {x_1}^2 - 2x_2y_2 + 2\alpha x_2 + 2\beta {x_2}^2$$

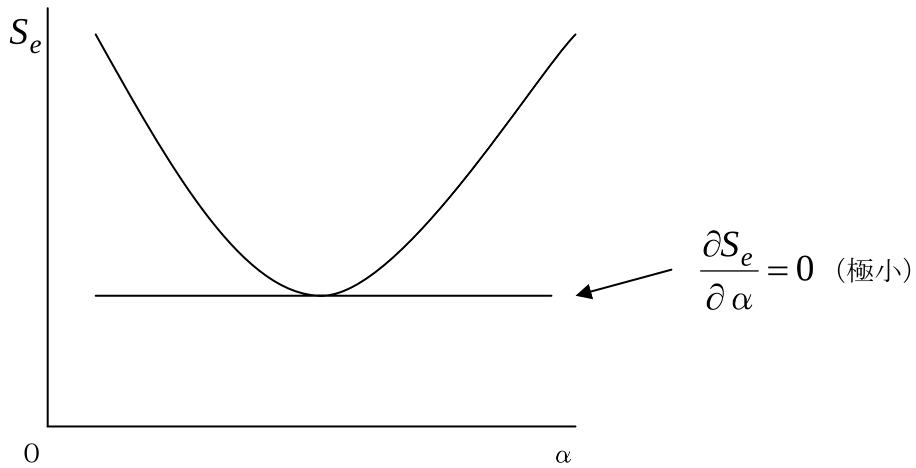
$$= -2(x_1y_1 + x_2y_2 - \alpha x_1 - \alpha x_2 - \beta {x_1}^2 - \beta {x_2}^2) = 0$$

$$\therefore x_1y_1 + x_2y_2 = \alpha(x_1 + x_2) + \beta({x_1}^2 + {x_2}^2) \quad \cdot\cdot\cdot [14]$$

→この[14]式は、前記正規方程式の[8]式 $\sum xy = \alpha \sum x + \beta \sum x^2$ に等しい。

---

【$S_e$ を α と β でそれぞれ偏微分してゼロとおくことの意味】




$S_e$ の値が極小となるための条件は、接線の傾きが０であること、すなわち１階の微分が０となることである（なお正確には、２階の微分が正（下に凸）であることが暗黙の前提である）。

以上